DOCUMENT RESUME

ED 284 895 TM 870 442

AUTHOR Terenzini, Patrick T.
TITLE The Case for Unobtrusive Measures.
PUB DATE 86
NOTE 17p.; In: Assessing the Outcomes of Higher Education; see TM 870 436.
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Colleges; Cost Effectiveness; *Educational Assessment; Error of Measurement; Evaluation Criteria; *Evaluation Methods; Higher Education; *Informal Assessment; *Institutional Evaluation; *Naturalistic Observation; *Outcomes of Education; Student Evaluation; Testing Problems
IDENTIFIERS *Unobtrusive Measures

ABSTRACT

Unobtrusive measures are recommended as a means of assessing educational outcomes of colleges. Such measures can counteract the response bias which is common in questionnaires and interviews. Outcomes researchers are, in fact, asked to supplement standard measures with unobtrusive measures. Interesting data may result from observation of students' social interaction and seating patterns, or the tenor of student graffiti and campus bulletin boards. Unobtrusive measures are warranted on the measurement grounds of reliability due to multiple measures, or increased sampling. In addition, observation of behavior is a more valid measure of attitudes than an interview in which the responses may be biases. Unobtrusive measures may be much less costly than standard tests, since conclusions may be drawn from data which have already been collected and stored. The prudence of unobtrusive measures is demonstrated by the wisdom of using multiple measures to assess something as multifaceted as the quality of a college education. Some examples of this type of measurement include the analysis of students' course-taking patterns. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284895

The Case for Unobtrusive Measures

Patrick T. Terenzini
Unikversity of Georgia

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Wedemeller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 870 448

Paper presented at the forty-seventh ETS Invitational Conference, sponsored by Educational Testing Service, at The Plaza, New York City, on October 25, 1986.

The Case for Unobtrusive Measures

PATRICK T. TERENZINI
University of Georgia

There can be little doubt that much of what we know from the social sciences has been developed from interview or questionnaire data. But what can we say about the fidelity of those portraits for the social and educational behaviors and phenomena they depict? Consider the following:

> ...research in intelligence testing show(s) that dependable gains in test-passing ability (can) be traced to experience with previous tests even where no knowledge of results (has) been provided...Similar gains have been shown in personal adjustment' scores (Webb et al., 1966, p. 19).

> Male interviewers obtain fewer responses than female, and fewest of all from males, while female interviewers obtain their highest responses from men, except for young women talking to young men (Benney, Riesman, & Starr, 1956, p. 143).

> Sequences of questions asked in very similar format produce stereotyped responses, such as a tendency to endorse the righthand or the lefthand response, or to alternate in some simple fashion. Furthermore, decreasing attention produces reliable biases from the order of item presentation (Webb et al., 1966, p.20).

Thus, much of what we know may be biased in various and sometimes unknown ways. But if what one blind man learns about elephants is biased by the data-gathering procedures adopted, measurement and sampling theory suggest it is reasonable to expect that the evidence gathered by multiple blind men, when pooled, will give a better, if imperfect, approximation of an elephant. There is, after all, more than one way of knowing. The central thesis of this paper is that multiple research designs and measures of educational outcomes are more likely to yield reliable and valid assessments of educational outcomes than is the current reliance on

interviews and questionnaires.

Consider the following:

The wear in the floor tiles in Chicago's Museum of Science and Industry.
The shrinking diameter of a circle of seated children.
Pupil dilation in the eyes of jade customers.
The bullfighter's beard.

Each of these conditions can, under certain circumstances, be taken as a measure of a phenomenon of interest to someone. Taken from Webb et al. (1966), each is an example of what has come to be called "unobtrusive measures," a general class of measurements presumed to reduce or eliminate the potential for reactive bias: responses uncharacteristic of the attitude or behavior outside the measurement situation and induced by the measurement act itself. The premise is that when interviews and questionnaires are used in social science research, the process of data collection intrudes itself into the consciousness of the subject and, as a consequence, alters the subject's responses. Unobtrusive measures, by their nature, avoid most, if not all, of the reactive bias associated with interview and questionnaire methodologies.

Webb et al. (1966) have described five categories of unobtrusive measures: physical traces (natural erosion or accretion processes, such as the wear on library book pages or the refuse left behind by an earlier civilization); continuous archival records (e.g., actuarial records, government records); intermittent archives (e.g., written documents, sales records); simple observations (e.g., of behaviors), and physical devices (e.g., cameras, video and audio tapes).

The measures listed above index some interesting illustrations of physical traces and simple observations. For example, the fact that the floor tiles around the hatching-chick exhibit require replacement approximately once every six weeks, compared to a replacement rate of several years for the tiles around other exhibits, can be taken as a reasonably clear reflection of the relative interest-value of the exhibit. So far as the shrinking diameter of the circle of children is concerned, if it were also known that the shrinkage was observed during a ghost-story-telling session, then the observation would have been recognized for what it is: an unobtrusive measure of the degree of fear induced in the children by the stories (and how much more reliable and valid than what the children might tell us if asked, "How scared were you?"). As for the dilation of the pupils in customers' eyes, Chinese jade dealers have used it as an indicator

of customer interest in various stones. And bullfighters' beards have been observed to be longer on days when the matador must enter the ring. There is no consensus whether the longer growth is attributable to higher anxiety or to whether he simply stands farther from the razor on those days. Probably both (Webb et al., 1966, pp. v and 2).

Much has been written on how the methods of the social sciences might be brought to bear on questions of outcomes assessment in higher education (e.g., Ewell, 1985, 1985; American College Testing Program, 1980; Astin, 1977). Less attention, however, has been given to the measurement problems inherent in these methods and to how those problems might be avoided or at least counterbalanced. Some critics consider the present reliance on interviews and questionnaires to be both unwise and unnecessary. Webb et al. (1963), for example:

> lament this overdependence upon a single, fallible method. Interviews and questionnaires intrude as a foreign element into the social setting they would describe, they create as well as measure attitudes, they elicit typical roles and responses, they are limited to those who are accessible and will cooperate, and the responses obtained are produced in part by dimensions of individual difference irrelevant to the topic at hand.
>
> *But the principal objection is that they are used alone* (p. 1; emphasis in the original).

Unobtrusive measures, such as those listed above, offer an important methodological counterweight to the unknown and unbalanced reactive bias in interview- and questionnaire-generated data sets, such as those upon which we now rely to assess the educational outcomes of college.

The remedy for these ailments, of course, lies not in the replacement of the research tools now in widespread use. This is no call to rally the Assessment Luddites. Rather, the intention is to encourage outcomes researchers to supplement standard approaches with methods and measures now largely unknown, unconsidered, or ignored. The purpose, here, is to make "The Case for Unobtrusive Measures," and that warrant can be argued on at least three grounds (one major, and two secondary): 1) measurement, 2) cost, and 3) prudence.

The Measurement Warrant

The strongest arguments for the use of unobtrusive measures can be made

(appropriately enough) on measurement grounds. Recall that the principal objection of Webb et al. (1966) to the current reliance on interviews and questionnaires was that "they are used alone" (p. 1). The foundation of this objection is that:

> Every measurement procedure carries with it certain characteristic sources of error...it follows that they are in error in different ways and different degrees. The errors we refer to are constant *within* types of measures—the direction and size of the error are assumed to be fixed for a given set of measurement operations. However, the direction of errors is assumed to be random *across* procedures. For any given measurement task, the errors are additive: an error in one direction will tend to cancel out an error in the other direction (Sechrest and Phillips, 1979, p. 2).

Sechrest and Phillips go on to note problems occasioned by differences in the magnitudes of the errors involved and their effects on the precision of measurement, but the point is clear and the strongest argument for the use of multiple and *different* measures of the same trait or behavior—what Webb et al. (1966) and others (e.g., Campbell & Fiske, 1959) refer to as "multiple operationism." The intent is to employ multiple measures that "share in the theoretically relevant components (of the trait or behavior under study) but have different patterns of relevant components" (Webb et al., 1966, p. 3). When one samples measures, one also samples their strengths *and* their weaknesses. And as in sampling theory, the larger the sample size, the greater the reliability of estimation.

The utility of multiple measures in general, and unobtrusive measures in particular, is apparent in another way. Much of the research on student outcomes, particularly that focusing on institutional contributions to student growth, relies on various causal modelling techniques based on multiple regression. The multicolinearity among theoretically independent predictor variables, and the autocorrelations among the same measures used over time in longitudinal designs, present well-known, but frequently ignored, problems for the interpretation of path coefficients or regression weights. The problems of "bouncing betas" and the difficulty of replicating most studies in the social sciences are also well-known. Such interpretive difficulties notwithstanding, however, one researcher (cited in Kerlinger & Pedhazur, 1970, p. 440) has suggested that regression coefficients give us the laws of science, and many who employ regression analysis, or who read and rely on the results of such studies, may be similarly inclined to place more credence in the findings than is warranted.

The wisdom of multiple—and unobtrusive—measures is evident in still other ways. Research on the dynamics of attitude and value formation and change has both perceptual and behavioral dimensions. What correspondence exists between what a respondent professes to believe and how that person actually behaves? Reliance on questionnaires and interviews in such investigations requires an act of faith that the correspondence is high, when the fact of the matter may very well be otherwise. One can have significantly greater confidence in the reliability and validity of interview- or questionnaire-based claims about attitudes and beliefs if those claims are manifested behaviorally in natural settings. Used in this fashion, unobtrusive measures constitute a form of convergent validation and go a long way toward reducing the internal validity problems inherent in *ex post facto* research designs.

Unobtrusive measures have their own limitations, of course, for we rarely, if ever, know *their* characteristic sources of error. Thus, we cannot confidently estimate the extent to which use of an unobtrusive measure would be a useful and complementary addition to a series of measurement procedures or simply increase the error already present. And, like interview and questionnaire items, to the extent that unobtrusive measures rely on single observations, they are likely to be unreliable and, consequently, of limited validity (Sechrest & Phillips, 1979, pp. 5-7). Despite more than a two-decade history, much research remains to be done on the measurement characteristics of unobtrusive measures.

Before all hope and confidence in the utility of unobtrusive measures is abandoned, however, it is useful, at least insofar as the assessment of educational outcomes is concerned, to differentiate "unobtrusive measures" as a set of scientific research tools from "unobtrusive measures" as a metaphor. In the first instance, it is quite possible to apply unobtrusive techniques and measures in a remarkable variety of experimental studies (see Bochner, 1979). As such, the rigor characteristic of true experiments can be brought to bear in naturalistic settings (like colleges and universities) and threats to internal validity are significantly reduced if not eliminated.

For example, if an institution wished to know the extent to which cultural and racial openness was a trait characteristic of the campus, one might design a study similar to that reported by Campbell, Kruskal and Wallace (1966). In that investigation, the tendency of White and Black college students to sit by themselves in racially homogeneous groups in classrooms (rather than mixing randomly) was studied as an indicator of racial attitudes.

While such formal, unobtrusive research efforts are certainly possible, they are probably not likely to comprise a complete or adequate outcomes assessment program. "Unobtrusive measures" as a metaphor for non-reactive sources of information that *already exist* in various forms and locations across a campus are more likely to yield useful vehicles of assessment. Examples include such standard records as registrar's files, disciplinary records, Graduate Record Examination (GRE) scores, and alumni giving records. The category can also include less conventional measures, however, ranging from transcripts sent to other undergraduate institutions (student satisfaction), to case loads in the health services and psychological counseling service (amount of stress on campus), to library usage rates (students' intellectual curiosity). Unobtrusive measures may be based on observations as well as records. Such measures in colleges and universities might include assessment of a campus's intellectual climate as revealed on bulletin boards and in graffiti (see Ciardi, 1970, for a delightful discussion on this topic) and in conversations overheard in a student union snack bar. The point to be made is that unobtrusive measures—whether scientifically formal or casual—offer a source of information about the educational process and its outcomes that serves a legitimate and important measurement role by counterbalancing the systematic error characteristic of conventional measurement and research designs and by validating information gathered by means of those standard procedures.

The Cost Warrant

The costs of assessing educational outcomes are little understood. The proponents of the "benefits" portion of the cost-benefits equation have been dominant, and only recently has attention been turned to an estimation of the other side of the balance scale. How much in the way of resources is and should be invested in the production of outcomes information? The question applies to all information gathering, of course, whether outcomes or otherwise, but costs in other sectors are better understood and estimated than they are in outcomes assessment. The real issue, as Ewell and Jones (1986) put it, is: "How much *more* money (beyond that already committed to outcomes-related information gathering) do we have to spend to put in place an assessment program that is appropriate to our needs?" (p. 34).

Based on a set of assumptions about the nature of the assessment

programs likely to be mounted by institutions of varying types and sizes, Ewell and Jones (1986) estimate incremental costs ranging from $30,000 (in a small, private, liberal arts college) to $130,000 (in a major public research university). It is important to bear in mind that these are incremental, not total, cost estimates. It is revealing to notice Ewell and Jones' assumption of the use of conventional questionnaires, whether commercially available (e.g., The ACT's COMP) or locally developed (e.g., senior examinations in the major field disciplines).

No one has attempted to estimate the incremental costs of assembling information unobtrusively. Given the fact that much of this sort of data already exists, and given that much of it is electronically stored and retrievable, it seems reasonable to suggest that the costs of unobtrusive measurement and analysis are likely to be lower than those of more conventional measures and methods, perhaps significantly lower. There is, of course, considerable room for cost variability, but the initial proposition holds: analyzing data that are already available in one form and place or another is likely to be less costly and time-consuming than gathering data *de novo*.

The Prudence Warrant

Ewell (1984) has written that "the most vehement objections to the systematic assessment of institutional impact will come from faculty" (p.72). These objections, says Ewell, are likely to derive from either or both of two sources: first, the fear of being negatively evaluated, and second, a philosophical opposition based on the belief that the outcomes of college are inherently unmeasurable and that the evidence from such studies is "misleading, oversimplifying, or inaccurate" (p. 73).

To counter faculty opposition, Ewell recommends that persons responsible for outcomes assessments "recognize publicly the inadequacy of any *single* outcome measure or indicator and . . . collect as many measures of program effectiveness as possible" (p. 73). The point is related to the argument for unobtrusive measures made earlier on measurement grounds and is likely to be recognized and given weight by faculty of all disciplines. The effect is likely to be a reduction in faculty resistance to educational assessment. Even if the measurements cannot be easily explained to non-social scientists, most faculty members will be familiar with the concept of "triangulation" in astronomy, as well as in map-reading and surveying. The use of multiple measures to portray some educa-

tional outcome is likely to have a face validity that is appealing to faculty members. It seems reasonable to expect such an effect to influence positively both faculty participation in outcomes assessment programs and confidence in the conclusions derived from the evidence assembled.

Unobtrusive Measures in Higher Education

What are some unobtrusive measures in higher education and how might they enhance our understanding of various educational outcomes? Ewell (1984), following a review of various structures and taxonomies, has suggested that educational assessment should focus on three major areas: knowledge, skills, and values and attitudes, with a fourth category, students' relations with various groups in the larger society, representing the behavioral manifestations of the first three areas. Juxtaposition of these four dimensions against three of the general types of unobtrusive measures described earlier affords a useful framework for thinking about the sorts of institutional information that might be used to aid educational assessment. The matrix below is intended to be suggestive, to focus thinking on important assessment topics, and, thereby, to highlight the potential opportunities to employ unobtrusive measures.

Types of Unobtrusive Measures

Outcome Categories	*Physical Traces*	*Archives & Records*	*Observations*
Knowledge			
Skills			
Attitudes/Values			
Relations w/Society			

Space precludes discussion of possible measures that might occupy each of the cells in this matrix, and, as will be seen, the boundaries between the several categories of unobtrusive measures are not always precise. Moreover, some of the cells are of greater interest than others, and some unobtrusive measures are more readily accessible than others. Two cells easily meet both of these criteria, namely, the "Knowledge-Archives" cell and the "Attitudes/Values-Observations" conjunction, and attention will be focused on them for illustrative purposes, beginning with the latter of the two cells.

The observational techniques of Campbell, Kruskal and Wallace (1966) for inferring racial attitudes and relations on a campus have been summarized. Variations on this approach might include a study of "aggregating" (Campbell, Kruskan and Wallace, 1966) in dining halls and cafeterias, in clusters of students studying in the library or gathering in other public areas, in institutional residence hall roommate patterns, and in other institutional settings.

Something of the importance students attach to the life of the mind might be inferred from several sources, including the number, size, and participation rates in formal student organizations and clubs that have some specific, academic purpose (e.g., discipline-based clubs, literary and artistic publications, performing arts groups), as compared with organizations that have athletic, recreational, entertainment, social, or other purposes as their principale *raison d'être*. (Some of this information might be gleaned from records.)

Similarly, inferences about the relative emphasis given to the academic and social life of a campus might be made based on an examination of the content of campus concert, film, lecture, and speaker series, as well as attendance records. For residential campuses, the institution's role in students' lives—and its potential for influence—may be reflected in the extent to which students evacuate the campus for other locations on weekends. Ciardi (1970) has suggested that the content of graffiti reflect the intellectual tenor of a campus. One might add the content of bulletin boards to that reflection.

The number of students who are registered—and active—voters can be taken as a sign of students' interests in, sense of responsibility toward, and willingness to participate in the political life of a larger community. Or one might explore the level of social responsibility in a student body by designing an experiment around the frequency with which students returned library books that were presumably "lost." More simply, the proportion of the library's total overdue volumes that are signed out to students (or faculty) provides at least one index of the level of simple courtesy, if not social responsibility, on a campus. Vandalism, both in absolute magnitude and rate of change over time, offers another reflection of the quality of life and the attitudes and values prevalent on a campus. As suggested earlier, the rates over time at which the health service's physicians prescribe stress-related medicines, and variations in the case loads of the counseling center staff, might both be used to index the amount of potentially unhealthy—and perhaps educationally dysfunctional—stress in the campus environment. Hodgkinson and Thelin

(1971) offer an impressive list of other possibilities. The variety is limited only by one's imagination and ingenuity.

Without question, the major impact of college on students' cognitive development is delivered through the curriculum, and any outcomes assessment program must deal in one fashion or another with the curriculum and with classroom-based learning. A variety of reactive measures have been developed to assess the nature and extent of students' cognitive growth (e.g., the ACT-COMP and Graduate Record Examinations subject tests), and these measures are typically used in "value-added" research designs of varying degrees of sophistication and validity.

Warren (1984) and Pascarella (1986) discuss some of the conceptual and methodological limitations of this approach to educational assessment, and those critiques need not be reiterated here. The point to be made is that something of the nature and extent of student learning can also be inferred from unobtrusive measures, from a data base that already exists and that has reasonable claims to reliability, namely, the registrar's file, which contains extensive information on the courses students have taken and the grades received.

Fincher (1984), recognizing the weaknesses and disadvantages of the grade-point average as a criterion of what has been learned, also marvels that "it works as well as it does" (p. 380). He writes:

> . . .the freshman GPA will often display scalar features that are quite remarkable: a tenacious arithmetic mean, a standard deviation of about one-half letter grade, and a range of five or more standard deviation units. More remarkable, perhaps, the freshman GPA appears to be more immune to contamination than separate course grades are, and it is a relatively independent criteria despite being a faulty one. In addition, the freshman GPA is relevant to such educational decisions as the dean's list, student probation and dismissal, the maintenance of athletic eligibility, the continuance of scholarships, etc. If not a completely adequate criterion of academic performance, the freshman GPA still serves many educational purposes (p. 380).

Wilson (1983) reports that admissions measures are essentially as valid for predicting long-term GPA as freshman year GPA. Because of this property, Fincher suggests, cumulative GPA may yet be a useful measure in educational assessment and worthy of analysis. It might, for example, be used as the criterion in regression models and covariance analyses in which pre-college academic aptitudes and achievements (and other poten-

tially confounding variables) have been controlled in a study of the residual variance in long-term GPA attributable to student effort and to instruction and student learning. Similarly, if pre-college predictors of academic performance are found to have high multiple correlations with actual college achievement, reasonable suspicions might be raised about the overlap of high school and college coursework (Fincher, 1984).

The registrar's files offer other possibilities. For example, an examination of the distribution of courses taken by size and type of instruction (e.g., lecture, seminar, lab, independent study) might prove extremely revealing of the nature of the formal educational process experienced by students (e.g., graduating seniors). How many opportunities were there for students in small numbers to study with a faculty member? Such a review might focus on students' first two years. Do large lecture sections dominate students' early contacts with faculty and collegiate instruction? What is the relative balance of opportunities for active vs. passive student participation in their own learning? While recognizing that "small" is not necessarily "better," most faculty and administrators would probably be concerned if students' opportunities for small-group instruction were rare.

Examination of the relative proportional distributions of students majoring *and graduating* in particular disciplines will tell something of the nature of the educational program being delivered, and comparisons of such distributions, both one with the other and each over time, will detect shifting emphases in what students are interested in and what the institution is providing. Similarly, student retention rates, both within and across majors, may yield useful information. While such rates must be interpreted with considerable care, rates occupying one or the other tail in the distribution suggest something about students' views of the education afforded in those programs. Precisely what an extremely low retention rate means may be open to dispute, but at the very least it calls attention to the need for further investigation.

Transcript analysis affords a more detailed examination of curricula structures and student course-taking patterns and brings one still closer to the substance of students' formal education. Using this technique, Blackburn et al. (1976) undertook a national study of changes in degree requirements between 1967 and 1974, exploring the amount, structure, and content of general education, and the structure and flexibility in selected major degree programs. They found, for example, that the typical baccalaureate degree recipient in 1974, compared to 1967, had taken about 22 per cent less coursework in general education.

Galambos et al. (1985), in a study of teacher education in the states

comprising the Southern Regional Education Board, used transcript analysis to compare the course-taking patterns of teacher education and arts and sciences degree graduates. They found that, on the average, teacher education graduates took proportionally fewer general education credits in all areas except the social sciences than did arts and sciences graduates. Their analyses also led them to conclude that "Given latitude, some students will ferret out the routes of least resistance to meet their general education requirements, and then pass the word on to others" (Galambos et al., 1985, p. 78). Such a finding on an individual campus is likely to be justifiable cause for a detailed—and important—review of general education courses and requirements.

The State University of New York at Albany used transcript analysis to test a belief prevalent among faculty and administrators that students were not gaining a "general education" because the only degree requirements were those of the major program; all other degree credits were electives. The analysis provided information of the average number and percentage of course credits taken by graduates of each academic department in each of some 20 content areas. Results indicated that, while students in certain major field areas were apparently avoiding certain content areas (e.g., natural and physical sciences, or foreign languages), the deviations from what most academicians might consider a "general education" were by no means so great as had been anticipated.

A variation on this approach is afforded by the following matrix (adapted from Blackburn et al., 1976, who also offer some useful classification rules):

	Per Cent of Courses Taken	
Type of Course	*Breadth*	*Depth*
Required		
Restricted		
Elective		
Full Elective		

Using this matrix, a computer-based analysis of the transcripts of all students, or of sub-groups of students (e.g., selected majors, transfers, freshmen), would afford several kinds of information. It would reveal something of the variety and depth of the course work to which students have been exposed during any given period of time in their college careers. In addition, it would suggest the relative control over students'

course-taking exercised by the institution and the major department, and one might expect considerable variation across departments even within the same institution. Blackburn et al. (1976) offer a useful variant of the above matrix that differentiates general education requirements and courses from those of the major field. If still another variation were adopted to take into account *when* the course work was taken (i.e., a matrix that has the same breadth and depth columns, but has for its rows the time dimension of course-taking, say, lower and upper division), information would be gained on whether students are taking "breadth" courses prior to the selection of a major, or later in their college years, perhaps *after* the major program requirements have already been satisfied. The timing issue is important to the educational purposes of "general education" requirements. Do the requirements exist to ensure that students have a broad exposure to the various disciplines and on the basis of which they can make a more informed selection of a major program? Or are the requirements intended primarily to ensure that students are exposed to a broad intellectual experience at *some* point before they graduate?

Warren (1984) has suggested the analysis might be taken a step further. One might be inclined to believe, for example, that such course-taking pattern analyses do not provide a sufficiently detailed portrait of students' academic experience, for such analyses tell nothing of what students have learned. Warren suggests that a reasonable approximation of what has been learned might be obtained by reviewing examination questions and major paper assignments in courses that recur in the pattern of requirements for general education or for a specific major field — whether those courses are elective or required. As Warren (1984, p. 13) notes: "No pre-enrollment, normative, or comparative information need complement it. The assertion is simply that Program X as typically completed by a known number of students produces the described learning." A certain amount of faith is required, of course — faith that examinations and paper assignments reflect course content and that a passing grade reflects the occurrence of learning above some threshold of acceptability.

It should be evident by now that researchers in higher education have a wide variety of research designs and measures upon which to draw in their efforts to assess the outcomes of a college education. Thus far, however, the record indicates a virtually exclusive reliance on a subset of those designs and methods. The purpose of this paper has been to suggest ways in which conventional methods of assembling information on student growth might be supplemented in ways that illuminate rather

than obscure. Webb et al. (1966, p. 34) put it succinctly:

> So long as we maintain, as social scientists, an approach to comparisons that considers compensating error and converging corroboration from individually contaminated (measures), there is no cause for concern. It is only when we naively place faith in a single measure that the massive problems of social science research vitiate the validity of our comparisons.

References

American College Testing Program. *College Outcomes Measurement Project (COMP): Summary Report of Research and Development 1976-80.* Iowa City, Iowa: American College Testing Program, 1980.

Astin, A. W. *Four Critical Years.* San Francisco: Jossey-Bass 1977.

Benney, M., D. Riesman, & S. Star. "Age and Sex in the Interview." *American Journal of Sociology,* 1956, *62,* 143-152.

Blackburn, R., E. Armstrong, C. Conrad, J. Didham, & I. McKune. *Changing Practices in Undergraduate Education.* Berkeley: Carnegie Council on Policy Studies in Higher Education, 1970.

Bochner, S. "Designing Unobtrusive Field Experiments in Social Psychology." In L. Sechrest (ed.), *Unobtrusive Measurement Today, New Directions for Methodology of Behavioral Science, No. 1.* San Francisco: Jossey-Bass, 1979.

Campbell, D. T., & D.W. Fiske. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin,* 1959, *56,* 81-105.

Campbell, D. T., W.H. Kruskal, & W. P. Wallace. "Seating Aggregation as an Index of Attitude." *Sociometry,* 1966, *29,* 1-15.

Ciardi, J. "Graffiti." *Saturday Review,* 53, May 16, 1970, 10ff.

Ewell, P. *Information on Student Outcomes: How to Get It and How to Use It.* Boulder, CO: National Center for Higher Education Management Systems, 1983.

Ewell, P. *The Self-Regarding Institution: Information for Excellence.* Boulder, CO: National Center for Higher Education Management Systems, 1984.

Ewell, P. (ed.). *Assessing Educational Outcomes, New Directions for Institutional Research, No. 47.* San Francisco: Jossey-Bass, 1985.

Ewell, P., & D. Jones. "The Costs of Assessment." In C. Adelman (ed.). *Assessment in American Higher Education: Issues and Contexts*. Washington: U.S. Office of Education, Office of Educational Research and Improvement, 1986. (U.S. Government Printing Office, Document No. OR 86-301).

Fincher, C., "Educational Quality and Measured Outcomes." *Research in Higher Education*, 1984, *20*, 379-382.

Galambos, E. C., L.M. Cornett, & H.D. Spitler. *An Analysis of Transcripts of Teachers and Arts and Sciences Graduates*. Atlanta: Southern Regional Education Board, 1985.

Hodgkinson, H., & J. Thelin. *Survey of the Applications and Uses of Unobtrusive Measures in Fields of Social Science*. Berkeley: University of California at Berkeley, Center for Research and Development in Higher Education.

Kerlinger, F.N., & E.J. Pedhazur. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston, 1973.

Pascarella, E.T. *Are Value-Added Analyses Valuable?* Paper presented to the Educational Testing Service Invitational Conference on "Assessing the Outcomes of Higher Education." New York, October 25, 1986.

Sechrest, L., & M. Phillips. "Unobtrusive Measures: An Overview." In L. Sechrest (ed.), *Unobtrusive Measurement Today, New Directions for Methodology of Behavioral Science, No. 1*. San Francisco: Jossey-Bass, 1979.

Warren, J. "The Blind Alley of Value Added." AAHE *Bulletin*. September 1984, 10-13.

Webb, E.J., D.I. Campbell, R.D. Schwartz, & L. Sechrest. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally, 1966.

Wilson, K.M. *A Review of Research on the Prediction of Academic Performance after the Freshman Year*. College Board Report No. 83-2. New York: The College Board, 1983.